GENERALIZED EXTREME VALUE REGRESSION FOR BINARY RESPONSE DATA: AN APPLICATION TO B2B ELECTRONIC PAYMENTS SYSTEM ADOPTION

By Xia Wang and Dipak K. Dey

National Institute of Statistical Sciences and University of Connecticut

In the information system research, a question of particular interest is to interpret and to predict the probability of a firm to adopt a new technology such that market promotions are targeted to only those firms that were more likely to adopt the technology. Typically, there exists significant difference between the observed number of "adopters" and "nonadopters," which is usually coded as binary response. A critical issue involved in modeling such binary response data is the appropriate choice of link functions in a regression model. In this paper we introduce a new flexible skewed link function for modeling binary response data based on the generalized extreme value (GEV) distribution. We show how the proposed GEV links provide more flexible and improved skewed link regression models than the existing skewed links, especially when dealing with imbalance between the observed number of 0's and 1's in a data. The flexibility of the proposed model is illustrated through simulated data sets and a billing data set of the electronic payments system adoption from a Fortune 100 company in 2005.

1. Introduction. During the past three decades of information system (IS) research, quite a rich but diverse body of theoretical and empirical work has accumulated on the adoption and diffusion of information technology (IT) based innovations. Technological advances and widespread uses of the internet are allowing businesses to automate a wide range of their business processes including payments. Electronic payments system (EPS) is an integrated process, in which payment data is sent and received electronically from accounts payable to accounts receivable without human intervention, and is a critical component of the information economy. Despite the tremendous benefits that EPS can offer, its adoption in business-to-business (B2B)

transactions remains a challenge. In 2004, checks represented 81 percent of the typical organizations' payments, according to the Association for Financial Professionals (AFP) 2004 survey. A more recent survey by AFP finds that, even though the payments are undergoing an unprecedented period of change because of the decline of the check in favor of electronic payments, a majority of B2B payments continue to be made by check. The 2007 AFP survey indicates that the typical organization is still making 74 percent of its B2B payments by check. Researchers have tested different proposed theories by showing the causal mechanism and identifying influential factors. A number of barriers have been identified that appear to obstruct the wider adoption of electronic payments [Stavins (2003); Chakravorti and Davis (2004)].

Though exploring association based on the past data is critical to understand the technology adoption behaviors, the ultimate objective of EPS adoption studies would be accurate prediction as the field directly aims at studying and solving practical problems. With no doubt, accurate prediction is a key concern for practitioners since it is anticipated future behavior that guides managerial action. The immediate motivation of Bapna et al. (2010)'s study, as quoted in their paper, was a problem asked by the company, "...how can we predict the likelihood of a firm to adopt EPS promoted by the company so that we can target EPS promotions to only those firms that were more likely to adopt EPS for their payments?" Models with good prediction accuracy are also needed by theorists in the field since accurate prediction can serve as a direct "reality check" for the relevance of the theoretical models, thus providing a base for theory building, comparing and improving [Shmueli and Koppius (2009)]. A good statistical model for EPS adoption study needs to provide not only good fit to current data but also good prediction for future observations.

Currently, the logistic regression model, with its convenient interpretation and implementation, has been routinely employed to estimate and predict the EPS adoption or other new technology adoption in the literature [Chau and Jim (2002); Bapna et al. (2010); Gupta and Chintagunta (1994); Kamakura and Russell (1989); Wedel and DeSarbo (1993)]. When the logistic regression model is employed, it is assumed that the response curve between the covariates and the probability is symmetric. This assumption may not always be true, and it may be severely violated when the number of observations in the two response categories are significantly different from each other. This unbalance is not uncommon in the IS field, since in many cases the adoption of a new technology may be a "rare event," which happens with only a small probability. Thus, the data in this area is usually complicated by the nature of the response variable being analyzed: significant difference between the number of firms adopting the technology and of those

not adopting the technology and high skewness in the response curve. Applying a nonflexible link function to the data with this special feature may result in link misspecification.

Consequences of link misspecification have been studied by a number of authors in the literature. In particular, for independent binary observations, Czado and Santner (1992) show that falsely assuming a logistic link leads to a substantial increase in the bias and mean squared error of the parameter estimates as well as the predicted probabilities, both asymptotically and in finite samples. Moreover, these undesirable effects are of greater magnitude when the misspecification involves skewness than when it involves kurtosis (or tail weight). Wu, Chen and Dey (2002) show that under certain conditions there exists linear relationships between the regression coefficients, though the choice of links is important for goodness of fit.

There has been considerable work done in allowing flexibility in the link models used for categorical response data. The most intuitive approach to guard against link misspecification is to embed the symmetric links into a wide parametric class of links. Aranda-Ordaz (1981) introduces two separate one-parameter models for symmetric and asymmetric departures from the logistic model. Guerrero and Johnson (1982) suggest a one-parameter Box–Cox transformation of the odds. Morgan (1983) presents a one-parameter cubic logistic model to model symmetric departures from the logistic curve. It is a first-order approximation of the symmetric Aranda-Ordaz model.

Stukel (1988) proposes a class of generalized logistic models for modeling binary data with a two-parameter family. Stukel's models are general, and several important and commonly used symmetric and asymmetric link models can be approximated by members of this family. However, in the presence of covariates, Stukel's models yield improper posterior distributions for many types of noninformative improper priors, including the improper uniform prior for the regression coefficients [Chen, Dey and Shao (1999)]. Using a latent variable approach of Albert and Chib (1993), Chen, Dey and Shao (1999) propose another class of skewed links, which can lead to proper posterior distributions for the regression parameters using standard improper priors. However, the model has the limitation that the intercept term is confounded with the skewness parameter. This problem was overcome in Kim, Chen and Dey (2008) by a class of generalized skewed t-link models (GT-link), though the constraint on the shape parameter δ as $0 < \delta \leq 1$ greatly reduces the possible range of skewness provided by this model.

To build an appropriate and extremely flexible model for the binary data and to overcome the constraint for the skewed generalized t-link models, we propose the generalized extreme value (GEV) distribution as a link function. We would first distinguish our GEV model from a line of research on the discrete choice behavior initiated by McFadden (1978). Although their model is also termed as generalized extreme value models, it is totally different

from the model we introduce here. In their definition, GEV distribution is a family of multivariate distribution functions whose marginal distribution is Type I extreme value distribution or Gumbel distribution [McFadden (1978)], which is a special case of the GEV distribution we use in (3.2) with the shape parameter $\xi \to 0$. Without a flexible shape parameter, their model does not incorporate the skewness of the response curve but mainly estimates the perceived similarity among alternatives [McFadden (1978)]. The advantage of the GEV link model we discuss here is that it incorporates a wide range of skewness with the unconstrained shape parameter. In fact, the complementary log–log (Cloglog) link, based on the Gumbel distribution as discussed in Section 3, is a special case of the proposed GEV link.

For model comparison, we use measures which have been suggested as a criterion when the goal is to select a model with best out-of-sample predictive power, including Deviance Information Criterion (DIC) [Spiegelhalter et al. (2002)], Bayesian Information Criterion (BIC) [Kass and Raftery (1995)] and marginal likelihood [Chib and Jeliazkov (2001); Chib (1995)]. These three measures are all in-sample predictive measures, which are computed using current data. Shmueli and Koppius (2009) suggest that it is more general to construct predictive accuracy measures based on out-of-sample metrics. Thus, we also calculate posterior predictive deviance based on the hold-out sample approach. Comparison of predictive performance also guards against overfitting. Overfitting is a concern when building a complicated and flexible model. Since an overfitting model accommodates more of random fluctuations instead of the underlying relationship, it will generally provide good fit to current data but have poor predictive accuracy for new data. Good prediction results provide evidence that overfitting has not occurred.

The rest of the paper is organized as follows. Section 2 describes the data set concerning the electronic system adoption of the customers of a Fortune 100 company, where the dominance of nonelectronic payment methods is particularly evident. Section 3 describes the GEV model and discusses its posterior propriety with Bayesian methodology. Section 4 demonstrates the flexibility of the GEV model by simulated data sets and the EPS data set described in Section 2. It concludes with a discussion in Section 5. The proof of the theorem and detailed results on simulated data examples are given in the Appendix. The computation for the GEV link described in this paper has been implemented in R. A sample code [Wang (2010)] is included as supplementary material.

2. The electronic payment system data set.   To illustrate how the proposed GEV model may flexibly be used to model and to predict the EPS adoption, we consider a billing data set from a Fortune 100 companies (hereafter called the "vendor"). It provides information on individual transactions with firms who have financed purchases of large commercial equipment with

TABLE 1
Summary statistic for continuous variables

Variable	Min	Mean	Max
# of employees	1	4142	355000
Payment amount	11.4	17351.0	9520298.4
PayDex	5	71.83	89

the vendor before 2005. With a large amount to pay for the expensive commercial equipment, firms usually finance the purchases with the vendor by lease or loan ranging from a few months to more than 20 years. This brings in subsequently recurring payments every month. The vendor provides all client firms a voluntary free service of Automated Clearing House (ACH) debits. Firms have the option to make their payments using ACH or using conventional payment methods (such as a check). The data contains payments choices of each transaction in the first quarter of 2005 as well as firm-specific and transaction-specific information, including firm size, payment amount from the last bill, credit risk in terms of PayDex, finance option, geographical regions and industries. Firm size is measured by the total number of employees. The firm's credit risk is based on a PayDex score, which is an indictor of a business' payment performance evaluated on the total number of payment experiences over the past year. Higher PayDex scores indicate better payment performance. For each transaction, the firm has a financing option between lease and loan. Region is a geographic variable, while industry is a measure of firm demographics. Since these variables are easy to observe and often associated with usage behavior, they have been constantly used in marketing literature [Bapna et al. (2010)]. EPS is the electronic payment methods adopted by the firm, with EPS $=0$ for ACH methods and EPS $=1$ for traditional methods such as check. We have a total of 15,175 observations, with EPS $=1$ for 10,784 observations. That is, more than 70% of the transactions used traditional methods (EPS $=1$).

Summary statistics for this data set are reported in Tables 1 and 2. It provides a group of firms of different sizes, from different industries and regions, and various credit risk levels with different transaction amounts. As seen in Table 2, the proportions of EPS $=1$, which are almost all higher than 50%, vary within the subgroups based on finance option, region or industry.

We focus on a reduced sample to show the motivation of our proposed model. The subset is selected from all transactions that involved Midwest firms in the manufacturing industry with lease as the financing option. This results in 1618 transactions. Out of these transactions, 1222 transactions have EPS $=1$. We model the probability of EPS $=1$ with the normalized logarithm of firm sizes as the only convariate. This type of model would

TABLE 2
Summary statistic for categorical variables

Variables	EPS $= 1$	Counts	% EPS $= 1$
Finance option			
Lease	7857	10300	76.28%
Loan	2927	4876	60%
Region			
Midwest	3107	4608	67.43%
Northeast	1466	2564	57.18%
South	3688	4621	79.81%
West	2523	3383	74.58%
Industry			
Agriculture	127	256	49.61%
Construction	982	1556	63.11%
Finance	308	416	74.04%
Manufacturing	3340	4959	67.35%
Mining	246	305	80.66%
Public administration	318	327	97.25%
Retail trade	396	69	84.44%
Services	1599	2127	75.18%
Transportation	2320	3253	71.32%
Wholesale trade	1148	1508	76.13%
EPS	10784	15175	71.06%

facilitate showing the response curves under different link models in graphs. Figure 1(a) shows the original data and the fitted probability of the logit, probit and Cloglog model, respectively. To view it more clearly, we collapse firm sizes into 7 categories as shown in Table 3. Also shown in Table 3 are the estimated number of EPS $= 1$ under different link models. Figure 1(b) shows the observed values and the fitted values under different link models. The positively skewed Cloglog link fits the data better than the two symmetric links, though there is still large discrepancy between the observed values and the estimated ones. This is because the skewness inherited in the data is much higher than that which can be provided by the Cloglog link, whose skewness is fixed as a constant. A link model that can automatically change its shape of the response curve based on the data would greatly improve the estimation and prediction power of the model.

3. Generalized extreme value link model. We first specify the notation we use through the paper. Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ denote an $n \times 1$ vector of n independent binary random variables. Also, let $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})'$ be a $k \times 1$ vector of covariates, $i = 1, \ldots, n$. Suppose $\mathbf{X}$ denotes the $n \times k$ design matrix with rows $\mathbf{x}_i'$, and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients. Assume that $y_i = 1$ with probability p_i and $y_i = 0$
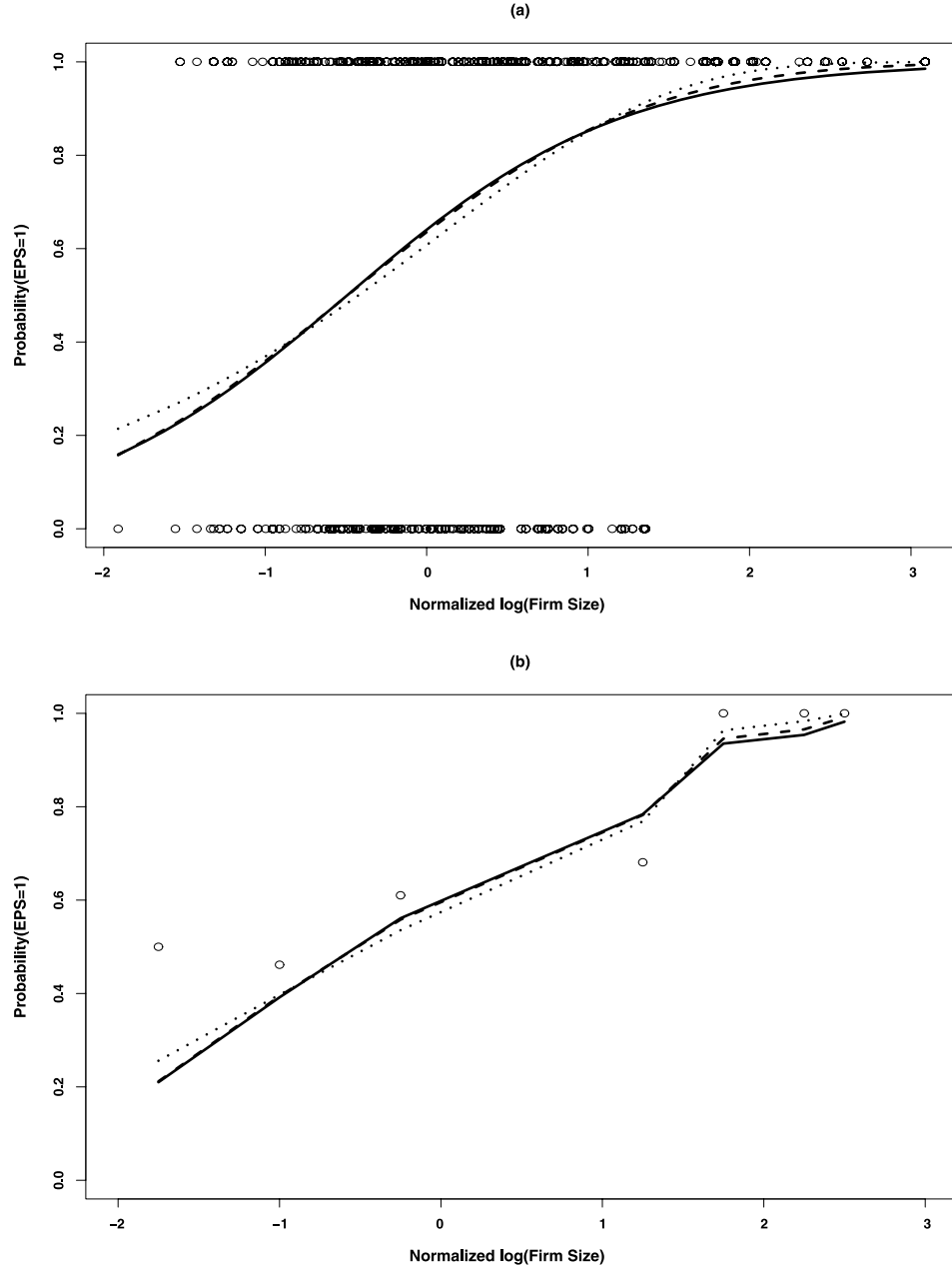
(a)

(b)

Fig. 1. *Model fitting using the logit (solid), probit (dashed) and Cloglog (dotted) link models to the subset of the EPS adoption data (circle) with the normalized logarithm of firm sizes as a covariate:* (a) *The original data;* (b) *collapse the firm sizes into 7 categories.*

TABLE 3
The estimated number of EPS $= 1$ in the transactions with Midwest manufacturing firms using lease as financing option under different link models

Normalized log(Firm size)	< -1.5	$(-1.5, -0.5)$	$(-0.5, 0)$	$(0, 1.5)$	$(1.5, 2)$	$(2, 2.5)$	> 2.5
# of EPS $= 1$	2	90	152	410	501	47	20
Logit	0.84	76.40	139.73	471.90	468.56	44.83	19.64
Probit	0.85	76.90	139.00	470.89	473.85	45.39	19.84
Cloglog	1.02	77.49	133.36	462.61	482.75	46.21	19.98

with probability $1 - p_i$. In the generalized linear model framework,

$$(3.1) \qquad p_i = \text{Probability}(y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}),$$

where F is a cumulative distribution function (cdf) and F^{-1} determines the link function. The function $F^{-1}(p_i) = \log\{p_i/(1 - p_i)\}$ gives the logit link, which is a symmetric link for binary response model. Another symmetric link model, the probit link, is achieved by setting $F^{-1}(p_i) = \Phi^{-1}(p_i)$, where Φ^{-1} is the inverse of $N(0,1)$ distribution. The asymmetric Cloglog link is specified as $F^{-1}(p_i) = -\log\{-\log(p_i)\}$.

3.1. *The generalized extreme value distribution.* The GEV link models use the Generalized Extreme Value (GEV) distribution for F. Extreme value theory begins with a sequence of independent and identically distributed random variables $Y_1, Y_2, \ldots$ and, for a given n asks about parametric models for its maximum $M_n = \max\{Y_1, \ldots, Y_n\}$. If the distribution of the Y_i is specified, the exact distribution of M_n is known. In the absence of such specification, extreme value theory considers the existence of $\lim_{n \to \infty} P[\{(M_n - b_n)/a_n\} \le y] \equiv F(y)$ for two sequences of real numbers $a_n > 0$ and b_n. If $F(y)$ is a nondegenerate distribution function, it belongs to either the Gumbel, the Fréchet or the Weibull class of distributions, which can all be usefully expressed under the umbrella of the GEV distribution with a cumulative distribution function as follows:

$$(3.2) \qquad G(x) = \exp\left[-\left\{1 + \xi\frac{(x - \mu)}{\sigma}\right\}_+^{-1/\xi}\right],$$

where $\mu \in R$ is the location parameter, $\sigma \in R^+$ is the scale parameter, $\xi \in R$ is the shape parameter and $x_+ = \max(x, 0)$. A more detailed discussion on the extreme value distributions can be found in Coles (2001) and Smith (2003). Extreme value analysis finds wide application in many areas, including climatology [Coles, Pericchi and Sisson (2003); Sang and Gelfand (2009)], environmental science [Smith (1989); Thompson et al. (2001)], financial strategy of risk management [Dahan and Mendelson (2001); Morales (2005)] and biomedical data processing [Roberts (2000)].
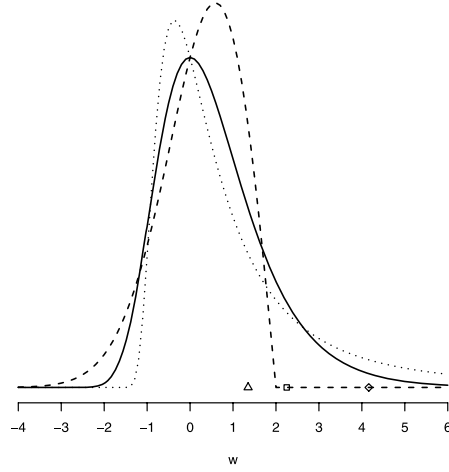
FIG. 2. *Probability density function plots of Weibull distribution ($\xi = -0.5$, dashed), Gumbel distribution ($\xi = 0$, solid) and Fréchet distribution ($\xi = 0.5$, dotted). The symbols $\triangle, \square$ and $\diamond$ are the 90th percentile of Weibull, Gumbel and Fréchet, respectively.*

Its importance as a link function arises from the fact that the shape parameter ξ in model (3.2) purely controls the tail behavior of the distribution (see Figure 2). When $\xi \to 0$, it gives the Gumbel distribution with $G(x) = \exp[-\exp\{-(x-\mu)/\sigma\}]$, which is the least positively skewed distribution in the GEV class when ξ is nonnegative.

3.2. *The generalized extreme value link model.* In model (3.1) we assume that F involves the GEV distribution as follows:

$$(3.3) \quad p_i = p(y_i = 1) = 1 - \exp\{(1 - \xi \mathbf{x}_i' \boldsymbol{\beta})_+^{-1/\xi}\} = 1 - \text{GEV}(-\mathbf{x}_i'\boldsymbol{\beta}; \xi),$$

where $\text{GEV}(x; \xi)$ represents the cumulative probability at x for the GEV distribution with $\mu = 0$, $\sigma = 1$, and an unknown shape parameter ξ.

Since the usual definition of skewness in (3.2) and (3.3) as $\mu_3 = \{E(x - \mu)^3\}\{E(x-\mu)^2\}^{-3/2}$ does not exist for large positive values of ξ's, we extend Arnold and Groeneveld (1995)'s skewness measure in terms of the mode to the GEV distribution for skewness definition and comparison. Under certain conditions, the skewness of a random variable X is defined as $\gamma_M = 1 - 2F(M_x)$, where $F(\cdot)$ is the cumulative distribution of X and M_x is the mode of X. Thus, the skewness of the link function (3.3) can be found explicitly as $\gamma_M = 1 - 2F(M_x) = 2\exp\{-(1+\xi)\} - 1$, while $\xi > -1$. Based on this skewness definition and on the fact that for $\xi \le -1$, $\partial p/\partial x$ monotonically decreases with respect to x in (3.3), we can show that the GEV link model specified in (3.3) is negatively skewed for $\xi < \log 2 - 1$, and positively skewed for $\xi > \log 2 - 1$.
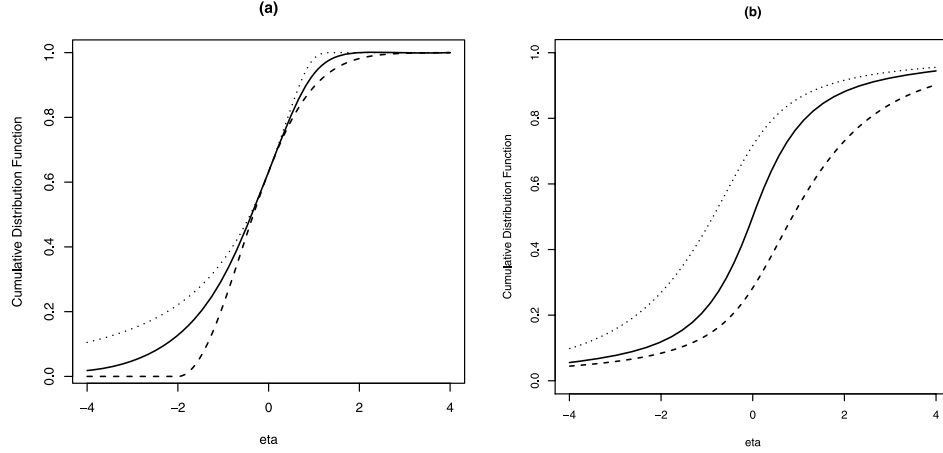
FIG. 3. *Cumulative distribution function plots of* (a) *the* GEV *link with* $\xi = -0.5$ *(dashed),* 0 *(solid) and* 0.5 *(dotted), and* (b) *the GT link of* $\delta z_i + \varepsilon_i$ *with* $v_1 = 1.2$, $\delta = 1/v_2 = 1$, *and* $G = \mathcal{E}$ *(dotted) and* $G = \mathcal{NE}$ *(dashed), and the solid curve is for* $G = \triangle_{(0)}$.

Figure 3(a) shows the response curves with ξ equal to $-0.5, 0$ and 0.5. The solid line is the response curve corresponding to the Cloglog link for $\xi \to 0$. As the values of the shape parameter change, so does the approaching rate to 1 and 0. A much wider range of skewness can be fitted compared to the commonly used Cloglog link. Figure 3(b) shows the distribution function of the skewed generalized t-distribution of $\delta z_i + \varepsilon_i$ with $v_1 = 1.2$, $\delta = 1/v_2 = 1$ [Kim, Chen and Dey (2008)]. The distribution G in the GT link is chosen as the standard exponential distribution $(\mathcal{E})$ and the negative standard exponential distribution $(\mathcal{NE})$. The skewness of the GT-link model is determined jointly by the constrained δ and the preassumed G. Since $\delta = 1$ is the maximum possible value for the shape parameter, Figure 3(b) in fact shows the two extreme cases of the GT-link with $G = \mathcal{E}$ or $\mathcal{NE}$, whose skewness is 0.4339 for the $GT(\mathcal{E})$ case and -0.4339 for the $GT(\mathcal{NE})$ case by Arnold and Groeneveld (1995)'s measure. The range of skewness provided by the GEV link models is not constrained. As shown in Figure 3(a), even with $\xi \in [-0.5, 0.5]$, the range of skewness provided by the GEV links is much wider than that provided by the GT links with a specified G.

The class of the GEV links also includes the symmetric link as a special case. For example, by matching the first 3 moments, the standard normal distribution can be approximated by the GEV distribution with $\mu \approx -0.35579, \sigma \approx 0.99903$, and $\xi \approx -0.27760$. Figure 4 shows the quantile plots between the GEV model and the probit model. The plot is approximately a straight line between 0.02 and 0.98 quantiles. The discrepancy lies mainly in the tail area.
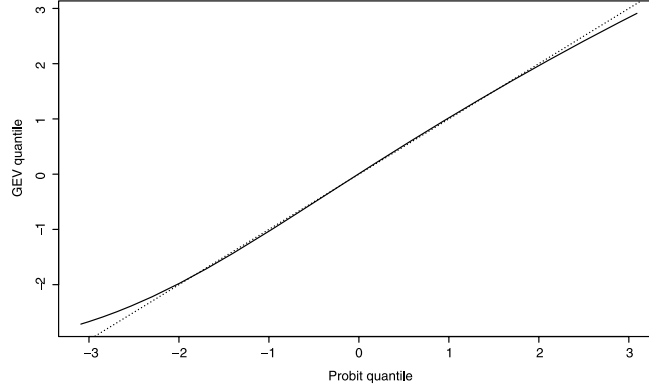
FIG. 4.   *Plot of* GEV *quantiles with* $\mu \approx -0.35579, \sigma \approx 0.99903$, *and* $\xi \approx -0.27760$ *against probit quantiles for probabilities between 0.001 and 0.999. The solid line is the quantile plot, and the dotted line is the* $45°$ *reference line.*

3.3. *Prior and posterior distributions for the generalized extreme value link model.* It is possible to estimate the shape parameter ξ in the GEV link by the maximum likelihood method. However, there are a number of nonregular situations associated with ξ in using likelihood methods which requires conditions for usual asymptotic properties associated with the maximum likelihood estimator to be valid. Smith (1985) studied the maximum likelihood estimation in nonregular cases in detail and obtained that when $\xi < -0.5$ the regularity conditions are not satisfied by GEV models. This violation of the usual regularity conditions means that the standard asymptotic likelihood results are not automatically applicable. This is one of the reasons for favoring a Bayesian analysis since Bayesian methods do not depend on the regularity assumptions required by the asymptotic theory of maximum likelihood. In particular, in the unusual situation where $\xi < -0.5$ and the classical theory of maximum likelihood breaks down, Bayesian inference provides a viable alternative. Thus, here we follow the Bayesian methodology and fit the GEV link model in a general setting.

Let $\mathbf{D}_{\mathrm{obs}} = (n, \mathbf{y}, \mathbf{X})$ denote the observed data. We assume that the prior of ξ, $\pi(\xi)$, is proper. Then the joint posterior distribution of $(\boldsymbol{\beta}, \xi)$ based on $\mathbf{D}_{\mathrm{obs}}$ is given by

$$(3.4) \qquad \pi(\boldsymbol{\beta}, \xi | \mathbf{D}_{\mathrm{obs}}) \propto p(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \xi) \pi(\boldsymbol{\beta} | \xi) \pi(\xi),$$

where $p(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \xi) = \prod_{i=1}^{n} \{1 - \mathrm{GEV}(-\mathbf{x}_i' \boldsymbol{\beta}; \xi)\}^{y_i} \{\mathrm{GEV}(-\mathbf{x}_i' \boldsymbol{\beta}; \xi)\}^{1-y_i}$ and $\pi(\boldsymbol{\beta} | \xi)$ is the conditional prior of $\boldsymbol{\beta}$ given ξ. In Sections 4.2–4.4 we choose normal priors for $\boldsymbol{\beta}$ with $\beta_j \sim N(0, \sigma_{\beta_j}^2)$, $j = 1, \ldots, k$, and for ξ with $\xi \sim N(0, \sigma_\xi^2)$, where priors on $\boldsymbol{\beta}$ and ξ are assumed independent.

Unlike Stukel's generalized logistic regression model with covariates, the posterior distributions under the GEV link are proper for many noninfor-

mative priors, including the Jeffreys' prior and the improper uniform prior for the regression coefficients $\boldsymbol{\beta}$.

Jeffreys' prior for this model has the form $\pi(\boldsymbol{\beta}|\xi) \propto |\mathbf{I}(\boldsymbol{\beta}|\xi)|^{1/2}$, where the Fisher information matrix $\mathbf{I}(\boldsymbol{\beta}|\xi)$ is $\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}$, with $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_n)$, $\omega_i = \{(1 - \xi\eta_i)^{-2/\xi - 2}\}[\exp\{(1 - \xi\eta_i)^{-1/\xi}\} - 1]^{-1}$, and $\eta_i = \mathbf{x}_i'\boldsymbol{\beta}$, for $i = 1, \ldots, n$. The joint posterior is then given by

$$\pi(\boldsymbol{\beta}, \xi | \mathbf{D}_{\mathrm{obs}}) \propto \prod_{i=1}^{n} \{1 - \mathrm{GEV}(-\mathbf{x}_i'\boldsymbol{\beta}; \xi)\}^{y_i} \{\mathrm{GEV}(-\mathbf{x}_i'\boldsymbol{\beta}; \xi)\}^{1-y_i} |\mathbf{I}(\boldsymbol{\beta}|\xi)|^{1/2} \pi(\xi).$$

The posterior propriety under the uniform prior established in Theorem 1 below implies that the proposed GEV link model is identifiable and the estimation of the regression coefficients can contain little subjective information in the Bayesian estimation. Let $\tau_i = -1$ if $y_i = 0$ and $\tau_i = 1$ if $y_i = 1$. Define $\mathbf{X}_{l,m}^* = (\tau_i \mathbf{x}_i', l < i \leq m)$ as the $(m - l) \times k$ matrix with rows $\tau_i \mathbf{x}_i'$, $l < i \leq m$, where $0 \leq l < m \leq n$. We are led to the following theorem concerning the propriety of the posterior distribution in (3.4) when $\pi(\boldsymbol{\beta}) \propto 1$, which is an improper uniform prior and $\pi(\xi) = 0.5$, for $-1 \leq \xi < 1$, which corresponds to the uniform distribution $U[-1, 1)$. Even with this constraint, the GEV links are still fairly flexible with skewness lying in the interval $[-0.7293, 1)$. The proof of Theorem 1 is given in the Appendix.

THEOREM 1. *Suppose that there exist $p > k$, $0 = m_0 < \cdots < m_p \leq n$, and positive vectors $\mathbf{a}_1, \ldots, \mathbf{a}_p$ such that $\mathbf{X}_{m_{l-1}, m_l}^*$ is of full rank and $\mathbf{a}_l' \mathbf{X}_{m_{l-1}, m_l}^* = \mathbf{0}$ for $l = 1, \ldots, p$ and with $\mathbf{0}$ a k-dimension zero vector. Under the improper uniform prior $\pi(\boldsymbol{\beta}) \propto 1$, the posterior (3.4) is proper.*

4. Applications of generalized extreme value link models.

4.1. *Model comparison criterion.* To assess models under different links, we employ four measures, including Deviance Information Criterion (DIC), marginal likelihood, Bayesian Information Criterion (BIC) and posterior predictive deviance methods. The first three can be deemed as model adequacy measures penalized by model complexity. Also, they measure the expected loss on replicated data and can be used as model predictive power measures. However, they are constructed based on current data [Shmueli and Koppius (2009)]. Posterior deviance measures the prediction accuracy directly based on hold-out samples. We briefly discuss their computation and interpretation below.

Let $\boldsymbol{\theta}$ denote the set of all parameters contained in the model under consideration. For example, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \xi)$ in the GEV model. The deviance is defined as -2 times the log-likelihood, that is, $D(\mathbf{y}, \boldsymbol{\theta}) = -2 \log p(\mathbf{y}|\boldsymbol{\theta})$. The posterior mean deviance $\hat{D}_{\mathrm{avg}}$ serves as a Bayesian measure of fit or "adequacy"

of a model [Spiegelhalter et al. (2002)], where $\hat{D}_{\mathrm{avg}}(\mathbf{y}) = \{\sum_{l=1}^{L} D(\mathbf{y}, \boldsymbol{\theta}_l)\}/L$ is the estimated average discrepancy for L sampling points and $\boldsymbol{\theta}_l$ is the lth sampling value for the parameter $\boldsymbol{\theta}$. The DIC measure, which is proposed by Spiegelhalter et al. (2002), is then calculated as $\mathrm{DIC} = \hat{D}_{\mathrm{avg}}(\mathbf{y}) + p_D$, where p_D is the effective number of parameters of a Bayesian model. We calculated p_D as $\hat{p} = \hat{D}_{\mathrm{avg}}(\mathbf{y}) - D_{\hat{\boldsymbol{\theta}}}(\mathbf{y})$, where $D_{\hat{\boldsymbol{\theta}}}(\mathbf{y}) = D(\mathbf{y}, \hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\theta}}$ is the posterior mean of the Markov chain Monte Carlo (MCMC) simulation. The smaller the DIC value, the better the model fits the data.

The marginal likelihood method is closely related to the Bayes factor, which is given by $m(\mathbf{y}|M_2)/m(\mathbf{y}|M_1)$ with $m(\mathbf{y}|M_i)$ the marginal likelihood under model M_i, $i = 1, 2$. The calculation of the marginal likelihoods can be obtained by the MCMC method as presented in Chib (1995) and Chib and Jeliazkov (2001). Chib and Jeliazkov (2001)'s approach is more relevant here since our models involve the Metropolis–Hastings algorithm.

The calculation of marginal likelihood is sensitive to the choice of priors [Kass and Raftery (1995)], which may bring in potential problems when the priors provide little information relative to the information contained in the data. Under this situation, BIC is usually applied as an approximation and it does not require evaluation of the prior distributions [Kass and Raftery (1995)]. For a model with k parameters and n observations, BIC is given by $\mathrm{BIC} = -2\log\{p(\mathbf{y}|\hat{\boldsymbol{\theta}})\} + k\log(n) = D_{\hat{\boldsymbol{\theta}}}(\mathbf{y}) + k\log(n)$.

To assess the predictive power of the model, we aim at evaluating the posterior predictive density $p(\tilde{\mathbf{y}}|\hat{\boldsymbol{\theta}})$, where $\tilde{\mathbf{y}}$ is the future response data (the holdout data set) and $\hat{\boldsymbol{\theta}}$ are the posterior mean of the MCMC sampling from the training data. The posterior predictive density can then be evaluated by posterior deviance ($\hat{D}_{\mathrm{post}}$) as $D(\tilde{\mathbf{y}}, \hat{\boldsymbol{\theta}}) = -2\log p(\tilde{\mathbf{y}}|\hat{\boldsymbol{\theta}})$.

4.2. *Simulated data examples.* In the simulated data examples, we consider data sets simulated from the Cloglog and probit regression models. Our primary aim is to show the flexibility of the GEV link in fitting the data generated from various models with unbalanced number of 0's and 1's. The true parameters are set such that the proportion of 1's in the simulated data sets is around 70%, similar to the proportions in the AFP surveys and the EPS data set. We perform Bayesian analysis for a given simulated data set and assess the models using criteria stated in Section 4.1.

To match data scenario close to the EPS data, we generate 5 covariates in our model, including the intercept. The types of covariates represent those that occurred in the real data. It includes one intercept ($\mathbf{x}_1$), one continuous covariate generated from a standard normal distribution ($\mathbf{x}_2$) and three discrete covariates. Among the three discrete covariates, two are dummies for nominal categorical data with 3 groups ($\mathbf{x}_3$ and $\mathbf{x}_4$) and the other is binary categorical data ($\mathbf{x}_5$). All covariates are generated for sample sizes

$n = 200, 1000$ and 5000, respectively. Then we generate two simulated data sets with n independent Bernoulli response variables, y_i, respectively from (a) the Cloglog regression model as Simulation $\sharp 1$, and (b) the probit regression model as Simulation $\sharp 2$. The linear components of all the above regression models are $\mathbf{x}_i'\boldsymbol{\beta} = \beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + x_{4i}\beta_4 + x_{5i}\beta_5$, where $i = 1, \ldots, n, \boldsymbol{\beta} = (0, 1, 1, 0.5, -0.5)$ for the Cloglog model and $(0, 1, 1, 1.25, -0.25)$ for the probit model. For all the link models, we employ the Metropolis–Hasting algorithm with normal jumping kernels for the MCMC sampling. The convergence of all results is examined by the Bayesian Output Analysis (BOA) package in R.

Results are discussed in detail in the Appendix (Simulated Data Examples). The GEV link performs better than the symmetric logit link even with small sample size ($n = 200$) under Simulation $\sharp 1$. The estimated ξ includes the true value 0 in its 95% high posterior density (HPD) intervals. With increasing sample size, the variance of ξ decreases, which provides more precise estimation of the shape parameter. In Simulation $\sharp 2$, the GEV link approximates the symmetric probit link well, especially with large sample size. The impact of sample size on fitting the GEV link model is also reflected by model selection based on BIC and marginal likelihood methods. These two criteria tend to select a simpler model and the GEV link beats the logit link only when $n = 5000$ based on these two criteria under Simulation $\sharp 2$.

By the simulated examples, we also emphasize that the number of 1's and 0's is only an indicator of the possible skewness in the response curve, though unbalance is not unusual even with a symmetric link model, such as the probit link we used in Simulation $\sharp 2$. As suggested in Chen, Dey and Shao (1999), many factors, such as the distribution of covariates, may affect the choice of links of a given data set. This complexity is exactly why we propose a flexible link function like the GEV link, such that the link function itself can automatically detect and fit the symmetry or asymmetry in the response curve along with parameter estimation. The probability of link misspecification is reduced compared to applying a link model with a fixed skewness parameter.

4.3. *The motivation subset of the electronic payments system data.* We first fit the GEV model to the subset we discussed in Section 2. Figure 5 shows (a) the fitted response curve and (b) the fitted probability for the 7 categories based on the firm's size. It is similar with Figure 1 but with the GEV link added. As shown in Figure 5, the response curve under the GEV link stretches significantly to fit the observed values. Table 4 shows the estimated number of EPS $= 1$, where we can see that the GEV model provides estimated values that are very close to the observed ones. Also, we carry out the cross validation analysis with randomly selected 10% of the data as the holdout part and the remaining data as the training part. The holdout
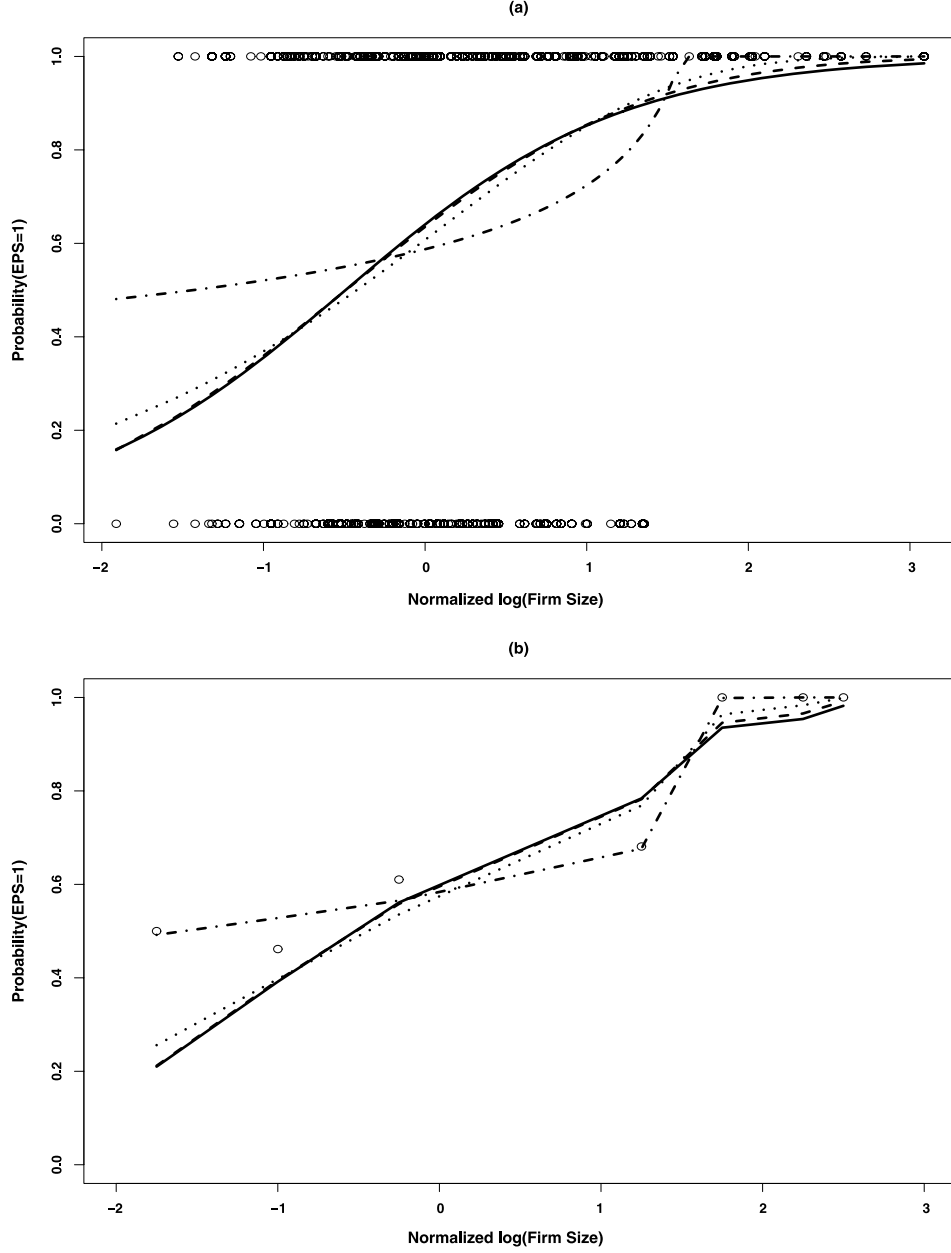
(a)

(b)

FIG. 5.   *Model fitting using the logit (solid), probit (dashed), Cloglog (dotted) and GEV (dotdash) link functions to the subset of the EPS adoption data (circle) with the normalized logarithm of firm sizes as a covariate. (a) The original data; (b) categorize firm sizes into 7 categories.*

TABLE 4
The estimated number of EPS $= 1$ in the transactions with Midwest manufacturing firms using lease as financing option under different link models

Standardized log(Firm size)	< -1.5	$(-1.5, -0.5)$	$(-0.5, 0)$	$(0, 1.5)$	$(1.5, 2)$	$(2, 2.5)$	> 2.5
# of EPS $= 1$	2	90	152	410	501	47	20
Logit	0.84	76.40	139.73	471.90	468.56	44.83	19.64
Probit	0.85	76.90	139.00	470.89	473.85	45.39	19.84
Cloglog	1.02	77.49	133.36	462.61	482.75	46.21	19.98
GEV	2	102.96	140.76	407.10	500.41	47.00	20.00

TABLE 5
The predicted number of EPS $= 1$ for transactions with Midwest manufacturing firms using lease as financing option under different links in the holdout part

Standardized log(Firm size)	< -1.5	$(-1.5, -0.5)$	$(-0.5, 0)$	$(0, 1.5)$	$(1.5, 2)$	$(2, 2.5)$	> 2.5
# of transactions	0	18	22	64	62	6	1
# of EPS $= 1$	—	5	19	42	62	6	1
Logit	—	7.20	12.20	49.71	57.93	5.74	0.97
Probit	—	7.23	12.13	49.57	58.58	5.81	0.98
Cloglog	—	7.23	11.63	48.64	59.73	5.91	1.00
GEV	—	9.37	12.25	42.92	61.94	6.00	1.00

part has 173 transactions with 135 transactions EPS $= 1$. The number of observed EPS $= 1$ in each of the 7 categories based on firm sizes is shown in Table 5. Here we can predict the number of EPS $= 1$ in each category using the posterior mean estimates obtained from the training part. It enables us to tell the predictive power of different models more directly than the posterior deviance measure. The GEV link still outperforms the other models and provides very good prediction for those firms with normalized logarithm sizes greater than 0.

4.4. *The electronic payment system data set.* To further illustrate the flexibility of the proposed GEV link models, we apply the model to analyze the whole EPS data. The aims of the analysis are to examine the effects of various factors on EPS adoption and to evaluate the fitness of regression models under different link functions. The continuous variables, firm size and the payment amount, are very skewed to the right. We take the logarithm and standardize them by the sample mean and standard deviation. The credit risk by Paydex is simply standardized. Financing option is a binary variable with 1 indicating that the firm uses lease for financing its purchase and 0 indicating loan. We have three dummy variables for the four regions,

TABLE 6

Model comparison under logit, probit, Cloglog and GEV links

Variables	Logit			Probit			Cloglog			GEV		
	est.	SD	ACE	est.	SD	ACE	est.	SD	ACE	est.	SD	ACE
Intercept	0.97	0.08		0.60	0.05		0.24	0.04		0.13	0.03	
Firm size	0.34	0.02	0.021	0.21	0.01	0.020	0.23	0.01	0.023	0.17	0.01	0.024
Payment amount	−0.17	0.02	−0.003	−0.11	0.01	−0.003	−0.11	0.01	−0.003	−0.06	0.01	−0.002
Credit risk	0.12	0.02	−0.029	0.07	0.01	−0.028	0.08	0.01	−0.034	0.07	0.01	−0.039
Financing option												
Lease	0.44	0.04	0.086	0.27	0.03	0.086	0.27	0.03	0.088	0.22	0.02	0.095
Region												
Mid west	−0.57	0.06	−0.109	−0.34	0.03	−0.109	−0.32	0.03	−0.103	−0.13	0.02	−0.059
North east	−0.85	0.06	−0.177	−0.53	0.03	−0.173	−0.54	0.04	−0.181	−0.42	0.03	−0.162
South	0.15	0.06	0.024	0.08	0.03	0.028	0.06	0.03	0.019	−0.01	0.01	−0.005
Industry type												
Agri., forest, fish	−1.15	0.14	−0.243	−0.70	0.09	−0.242	−0.80	0.10	−0.261	−0.99	0.18	−0.272
Consturction	−0.16	0.08	−0.029	−0.09	0.05	−0.030	−0.10	0.05	−0.033	−0.17	0.03	−0.071
Fin., ins., reale.	−0.07	0.13	−0.011	−0.03	0.08	−0.014	−0.04	0.07	−0.012	−0.05	0.04	−0.023
Manufacturing	−0.25	0.07	−0.044	−0.14	0.04	−0.047	−0.13	0.04	−0.041	−0.08	0.02	−0.034
Mining	0.42	0.16	0.057	0.20	0.09	0.070	0.10	0.08	0.031	−0.06	0.04	−0.024
Public adm.	2.58	0.36	0.255	1.38	0.16	0.257	1.10	0.12	0.248	0.48	0.05	0.194
Retail trade	0.43	0.15	0.069	0.24	0.08	0.073	0.18	0.07	0.054	0.02	0.03	0.007
Service	0.25	0.08	0.045	0.15	0.05	0.045	0.13	0.04	0.041	0.04	0.02	0.017
Trans., ele., gas	0.14	0.07	0.027	0.09	0.05	0.025	0.09	0.04	0.028	0.01	0.02	0.003
Shape parameter (ξ)										1.40	0.08	
Marginal likelihood	−8489.2			−8485.2			−8450.14			−8359.24		
BIC-16000	888.71			862.63			790.64			585.38		
DIC-16000	759.45			732.84			660.79			446.68		
$\hat{p}$	17.21			16.94			16.91			17.30		

and 9 dummy variables for the 10 industries, taking the values 0 and 1. Thus, there are a total of 16 covariates in the data for the response EPS.

Table 6 shows the parameter estimation, its standard deviation (SD), the average covariate effects (ACE), marginal likelihood, BIC-16000, DIC-16000 and $\hat{p}$ for different link functions using the normal priors. The prior variances for $\boldsymbol{\beta}$'s and ξ are equal to 10^4. A factor of 10 changes in these variance settings led to almost identical posterior results.

We obtain $\hat{\xi} = 1.40$ with a standard deviation of 0.079 for the GEV link model, which indicates that ξ is significantly above 0. In fact, the value of DIC-16000 of the GEV link model is 446.68 with the effective dimension $\hat{p} = 17.30$, which is lower than 660.79 of the Cloglog link model with $\hat{p} = 16.91$. Both skewed link models are better than the symmetric logit and probit link models. The consistent model comparison results are obtained by using the marginal likelihood and BIC criteria.

A closer look at the results from the widely used logit link regression model in the IS research and our GEV regression model reveals some difference in the estimation of the covariates' effects on EPS adoption. The changes in covariates include doubling the firm size, doubling the payment amount, 10 points decrease in the paydex measure, or moving from 0 to 1 for all the discrete covariates. The average covariate effects are calculated based on Chib and Jeliazkov (2006). A point to clarify is that the changes in the firm size, payment amount and the credit risk are on their original levels, while the regression itself has been run with these covariates standardized. The logit model suggests that transactions by firms in the southern region may have a lower probability employing EPS compared to those by firms in the western region with 1.9–2.8% changes in the probability, but the estimation from the GEV model implies that there is no significant difference in these two regions with covariate effect approximately equal to a 0.54% increase in the probability. Also, the GEV model indicates that transactions by firms in the construction industry are significantly more likely to use EPS (7% higher in probability) compared with those by firms in the wholesale trade industry, while the logit model suggests that it is not a statistically significant factor with the estimated average covariate effects around 3%. Even when a parameter is significant in both models, the impact of changes in the covariate may be quite different. For example, firms in the Midwest are approximately 11% more likely to use check compared to those in the west based on the logit model, while the probability is only around 6% as estimated by the GEV model.

To calculate the posterior deviance of the four different link models, we randomly divide the data into training and hold-out parts, with 10% or 1501 observations as the hold-out part and we calculate the posterior deviance ($\hat{D}_{\text{post}}$), which are 1653.40, 1643.04, 1637.92 and 1618.73 for logit, probit, Cloglog and GEV links, respectively. The GEV link model, with the lowest $\hat{D}_{\text{post}}$, outperforms the other three commonly used models in prediction. Thus, there is no indication of overfitting in the proposed model.

Since the assignment of 0 and 1 to the response variable is arbitrary, we can reverse the role of 0 and 1. It does not affect the fitness of the symmetric links, but it changes the asymmetry in the data. If we now define EPS $= 1$ if the transaction used the ACH service and 0 if it used the conventional payment method, we expect that a positively skewed link model like the Cloglog link is not appropriate anymore. In fact, we fitted the data using the same MCMC sampling method, and the values of DIC-16000 for the logit, the probit, the Cloglog and the GEV link are 759.03, 733.1, 815.49 and 516.66, respectively. The Cloglog fits the data poorly, just as expected, since it is a positively skewed link function and the response function is now negatively skewed. This result shows that the determination of links should not just be a matter of choosing between a symmetric link and a skewed link,

and that the direction of a skewed link indeed plays a more important role, since a wrong choice of direction gives an even worse fit than a wrong choice between symmetric and asymmetric. The flexibility of the GEV link models is apparent here as it includes the shape parameter as a model parameter to be estimated through the data, instead of preassuming any direction of it. With the redefined 0 and 1 in the response variable, the GEV link model has a negatively significant shape parameter of -1.11 with a standard deviation of 0.05.

5. Conclusion and discussion. This paper introduced a new flexible skewed link model for analyzing binary response data with covariates in the EPS adoption setup. The proposed GEV link model provided great flexibility in fitting skewness in the response curve. Although theoretically the GT link models in Kim, Chen and Dey (2008) would be rich enough to have similar flexibility, it is not easily implementable from a computational perspective. Moreover, the computation burden is much less for the GEV link. For a simulated data with 2 covariates, it took approximately 2 minutes to fit the GEV model in R, while it took approximately 8 hours to fit the GT model with codes in FORTRAN 95.

One existing problem in the GEV link is that the shape parameter ξ also affects the heaviness of the tail in the distribution. Its flexibility would be further improved if we can design a mechanism to modify the GEV distribution such that one parameter would purely serve as skewness parameter while the other could purely control the heaviness of the tails.

APPENDIX

Proof of Theorem 1. Let $u, u_1, \ldots, u_n$ be independent random variables with common distribution function F, which is a GEV distribution with $\mu = 0$, $\sigma = 1$, and a shape parameter ξ. For $0 < a < 1$, it can be shown that $E|u|^a < \infty$ for $-1 \leq \xi < 1$. Observing that $1 - F(-x) = EI(u > -x)$ and $F(-x) = EI\{-u \leq -(-x)\}$, here I is an indicator function. Now, we have $\{1 - F(-\mathbf{x}_i'\boldsymbol{\beta})\}^{y_i}\{F(-\mathbf{x}_i'\boldsymbol{\beta})\}^{1-y_i} \leq EI\{\tau_i u_i \geq \tau_i(-\mathbf{x}_i'\boldsymbol{\beta})\}$ and $\{1 - F(-\mathbf{x}_i'\boldsymbol{\beta})\}^{y_i} \times \{F(-\mathbf{x}_i'\boldsymbol{\beta})\}^{1-y_i} \geq EI\{\tau_i u_i > \tau_i(-\mathbf{x}_i'\boldsymbol{\beta})\}$. Let $\mathbf{u}^* = (\tau_1 u_1, \ldots, \tau_n u_n)$. Using Fubini's theorem, we obtain

$$\int_{-1}^{1}\int_{R_k} p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \xi)\, d\boldsymbol{\beta}\, d\xi$$

$$\text{(A.1)} \qquad = \int_{-1}^{1}\int_{R_n} E\left[\int_{R_k} I\{-\tau_i \mathbf{x}_i'\boldsymbol{\beta} < \tau_i u_i, 1 \leq i \leq n\}\, d\boldsymbol{\beta}\right] dF(\mathbf{u})\, d\xi$$

$$= \int_{-1}^{1}\int_{R_n} E\left\{\int_{R_k} I(\mathbf{X}^*\boldsymbol{\beta} < \mathbf{u}^*)\, d\boldsymbol{\beta}\right\} dF(\mathbf{u})\, d\xi.$$

Under the condition of Theorem 1, it follows directly from Lemma 4.1 of Chen and Shao (2000) that there exists a constant K such that $\|\boldsymbol{\beta}\| \leq K \min_{1 \leq l \leq p}(\max_{m_{l-1} < i \leq m_l} |w_i|)$ whenever $\mathbf{X}^* \boldsymbol{\beta} \leq \mathbf{w}$, where $\mathbf{w} = (w_1, \ldots, w_n)$. Hence, from (A.1), we have

$$\int_{-1}^{1} \int_{R_k} p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \xi) \, d\boldsymbol{\beta} \, d\xi$$

$$\leq K \int_{-1}^{1} \int_{R_n} \prod_{l=1}^{k} E\left(\max_{m_{l-1} < i \leq m_l} |\mathbf{u}^*|^{k/p}\right) dF(\mathbf{u}) \, d\xi < \infty.$$

Simulated data examples. We fit Cloglog, GEV and logit models for Simulation ♯1. The GEV model gives almost identical estimates of the regression coefficient $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ as the true Cloglog regression model. The estimated shape parameter and its 95% high posterior density intervals are $0.402(-0.199, 1.193), 0.058(-0.269, 0.175)$ and $-0.034(-0.136, 0.073)$ as $n = 200, 1000$ and 5000. The true value of ξ is already contained in the 95% HPD interval when $n = 200$. However, the variance in the estimation of the ξ becomes smaller as the sample size increases, which implies that the GEV links do require enough information contained in the data set for more efficient estimation of the shape parameter.

Czado and Santner (1992) emphasized that it is more appropriate to study the link misspecificaiton effects on the estimated probabilities since estimation of $\boldsymbol{\beta}$ heavily depends on the chosen link function. They suggested that different link models can only be *unambiguously* compared in their estimation of event probabilities. Table 7 shows the average covariate effects. These effects are calculated by the method suggested in Chib and Jeliazkov (2006), where they marginalize out the covariates as a Monte Carlo average using their empirical distributions and integrate out the parameter $\boldsymbol{\theta}$ by their posterior distribution. The third column in Table 7 indicates that the average covariate effect is measured as $\mathbf{x}_2$ doubled, or $\mathbf{x}_3, \mathbf{x}_4$ or $\mathbf{x}_5$ moving from 0 to 1. The values from the Cloglog model are in bold font. The average covariate effects estimated by the GEV model are closer to those estimated by the Cloglog model than the logit model.

Table 8 shows the model comparison results. We also include the effective number of parameters to show model complexity. To test the predictive power of different link models, for each simulated data set, we generate another data set of the same size $n \, (= 200, 1000, 5000)$ with the same $\boldsymbol{\beta}$'s as the hold-out part and the originally simulated data set as the training part. Based on the deviance in Table 8, the GEV link outperforms both the Cloglog model and the logit model. The DIC measure allows us to take the model complexity into account. The DIC values for the GEV model are very

TABLE 7

Simulation ♯1 based on the Cloglog regression model: average covariate effect comparison

Sample size	Parameter	Covariate change	Cloglog	GEV	Logit
$n = 200$	β_2	double	**0.2071**	0.2033	0.1985
	β_3	$0 \to 1$	**0.2409**	0.2533	0.2191
	β_4	$0 \to 1$	**0.1100**	0.1150	0.0894
	β_5	$0 \to 1$	**−0.1332**	−0.1428	−0.1300
$n = 1000$	β_2	double	**0.1902**	0.1889	0.1826
	β_3	$0 \to 1$	**0.1771**	0.1737	0.1624
	β_4	$0 \to 1$	**0.0861**	0.0844	0.0807
	β_5	$0 \to 1$	**−0.1306**	−0.1330	−0.1460
$n = 5000$	β_2	double	**0.1925**	0.1924	0.1875
	β_3	$0 \to 1$	**0.2082**	0.2096	0.2152
	β_4	$0 \to 1$	**0.1016**	0.1021	0.1024
	β_5	$0 \to 1$	**−0.1122**	−0.1115	−0.1067

TABLE 8

Simulation ♯1 based on the Cloglog regression model: model comparison

		Cloglog	GEV	Logit
$n = 200$	$D_{\hat{\boldsymbol{\theta}}}$	176.57	175.64	181.38
	DIC	186.24	186.98	191.47
	$\hat{p}$	4.84	5.67	5.05
	marginal likelihood	−119.34	−120.96	−118.31
	BIC	203.06	207.43	207.87
	$\hat{D}_{\text{post}}$	168.65	169.13	173.34
$n = 1000$	$D_{\hat{\boldsymbol{\theta}}}$	890.96	890.69	896.78
	DIC	901.06	902.62	906.99
	$\hat{p}$	5.05	5.96	5.10
	marginal likelihood	−481.02	−486.37	−481.17
	BIC	925.50	932.14	932.32
	$\hat{D}_{\text{post}}$	863.69	863.81	872.52
$n = 5000$	$D_{\hat{\boldsymbol{\theta}}}$	4388.53	4388.19	4427.16
	DIC	4398.52	4399.88	4436.85
	$\hat{p}$	5.00	5.85	4.84
	marginal likelihood	−2233.73	−2240.97	−2250.31
	BIC	4431.11	4439.29	4469.75
	$\hat{D}_{\text{post}}$	4418.96	4419.69	4473.86

TABLE 9

Simulation ♯2 based on the probit regression model: average covariate effect comparison

Sample size	Parameter	Covariate change	Probit	Logit	Cloglog	GEV
$n = 200$	β_2	double	**0.1757**	0.1752	0.1757	0.1727
	β_3	$0 \to 1$	**0.1619**	0.1636	0.15443	0.1579
	β_4	$0 \to 1$	**0.2867**	0.2861	0.2871	0.2821
	β_5	$0 \to 1$	**−0.0328**	−0.03478	−0.0123	−0.0285
$n = 1000$	β_2	double	**0.1892**	0.1882	0.1925	0.1906
	β_3	$0 \to 1$	**0.1756**	0.1766	0.1838	0.1778
	β_4	$0 \to 1$	**0.2630**	0.2641	0.2512	0.2583
	β_5	$0 \to 1$	**−0.0683**	−0.0650	−0.0689	−0.0703
$n = 5000$	β_2	double	**0.1871**	0.1862	0.1866	0.1877
	β_3	$0 \to 1$	**0.2254**	0.2253	0.2253	0.2259
	β_4	$0 \to 1$	**0.2859**	0.2858	0.2825	0.2858
	β_5	$0 \to 1$	**−0.0420**	−0.0417	−0.0403	−0.0417

close to the Cloglog model and both the GEV model and the Cloglog model provide better fit than the logit model based on the DIC measure. This result is consistent with the fact that the Cloglog model is a special case of the GEV model. Simulation ♯1 is in fact based on a GEV regression model with $\xi \to 0$. The comparison is consistent also in the aspect of the predictive power. In Table 8 the GEV model has almost the same posterior deviance as the Cloglog model, while the logit model is worse than both models for different sample sizes. Comparison based on the marginal likelihood and BIC is only consistent with the above results when $n = 5000$, which may suggest that benefits applying a GEV model are more prominent with large sample size. The better predictive performance of the GEV model suggests that there is no overfitting by the proposed model.

The results from Simulation ♯2 are summarized in Tables 9 and 10. The main objective of Simulation ♯2 is to show that the GEV model can approximate a symmetric probit link model. The 95% HPD intervals of the estimated shape parameter ξ are $(-0.8, 0.17), (-0.36, -0.03)$ and $(-0.31, -0.13)$ for $n = 200, 1000$ and 5000, which all include the value $\xi \approx -0.27760$ at which the GEV distribution approximates the standard normal distribution. A more precise estimation of ξ does require larger sample size. The sample size also has impact on how the GEV model compares with the other models. As shown in Table 10, only at $n = 5000$ all the model comparison criteria except the marginal likelihood method agree that the more complex GEV link model exhibits some advantages compared to the logit

TABLE 10
Simulation ♯2 based on the probit regression model: model comparison

		Probit	Logit	Cloglog	GEV
$n = 200$	$D_{\hat{\boldsymbol{\theta}}}$	154.74	155.13	156.25	154.63
	DIC	164.56	164.90	166.10	165.96
	$\hat{p}$	4.91	4.89	4.92	5.67
	marginal likelihood	-108.23	-105.73	-109.25	-114.48
	BIC	181.23	181.62	182.74	186.42
	$\hat{D}_{\mathrm{post}}$	199.80	200.45	201.69	199.67
$n = 1000$	$D_{\hat{\boldsymbol{\theta}}}$	861.44	862.70	865.28	861.02
	DIC	871.17	872.64	875.31	872.68
	$\hat{p}$	4.86	4.97	5.01	5.83
	marginal likelihood	-466.17	-463.97	-468.05	-472.90
	BIC	895.98	897.23	899.82	902.46
	$\hat{D}_{\mathrm{post}}$	843.24	844.42	851.27	844.60
$n = 5000$	$D_{\hat{\boldsymbol{\theta}}}$	4219.46	4226.72	4238.51	4217.05
	DIC	4229.43	4236.51	4248.40	4228.71
	$\hat{p}$	4.99	4.89	4.94	5.83
	marginal likelihood	-2149.16	-2149.96	-2158.67	-2155.75
	BIC	4262.04	4269.31	4281.09	4268.15
	$\hat{D}_{\mathrm{post}}$	4190.66	4196.77	4228.71	4195.57

and Cloglog links. The marginal likelihood method is affected by the flat priors on parameters. It requires an even larger sample size to select the GEV link as a better model.

SUPPLEMENTARY MATERIAL

R codes for GEV models with covariates (DOI: 10.1214/10-AOAS354SUPP; .txt). The computation for the GEV link described in this paper has been implemented in R which is available in this supplementary material.

REFERENCES

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. MR1224394

Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data. *Biometrika* **68** 357–364. MR0626394

Arnold, B. and Groeneveld, R. (1995). Measuring skewness with respect to the mode. *Amer. Statist.* **49** 34–38. MR1341197

Bapna, R., Goes, P., Wei, K. K. and Zhang, Z. (2010). A finite mixture logit model to segment and predict electronic payments system adoption. *Information Systems Research* DOI: 10.1287/isre.1090.0277.

Chakravorti, S. and Davis, E. (2004). An electronic supply chain: Will payments follow. *Chicago Federal Letter* **206a**.

Chau, P. and Jim, C. (2002). Adoption of electronic data interchange in small and medium-sized enterprises. *Journal of Global Information Management* **10** 61–86.

Chen, M.-H., Dey, D. K. and Shao, Q.-M. (1999). A new skewed link model for dichotomous quantal response data. *J. Amer. Statist. Assoc.* **94** 1172–1186. MR1731481

Chen, M.-H. and Shao, Q.-M. (2000). Propriety of posterior distribution for dichotomous quantal response models with general link functions. *Proc. Amer. Math. Soc.* **129** 293–302. MR1694452

Chib, S. (1995). Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc.* **90** 1313–1321. MR1379473

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *J. Amer. Statist. Assoc.* **96** 270–281. MR1952737

Chib, S. and Jeliazkov, I. (2006). Inference in semiparameteric dynamic models for binary longitudinal data. *J. Amer. Statist. Assoc.* **101** 685–700. MR2256181

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, New York. MR1932132

Coles, S., Pericchi, L. R. and Sisson, S. (2003). A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology* **273** 35–50.

Czado, C. and Santner, T. J. (1992). The effect of link misspecification on binary regression inference. *J. Statist. Plann. Inference* **33** 213–231. MR1190622

Dahan, E. and Mendelson, H. (2001). An extreme-value model of concept testing. *Management Science* **47** 102–116.

Guerrero, V. M. and Johnson, R. A. (1982). Use of the Box–Cox transformation with binary response models. *Biometrika* **69** 309–314. MR0671968

Gupta, S. and Chintagunta, P. K. (1994). On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research* **31** 128–136.

Kamakura, W. A. and Russell, G. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* **26** 379–390.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90** 773–795.

Kim, S., Chen, M.-H. and Dey, D. K. (2008). Flexible generalized t-link models for binary response data. *Biometrika* **95** 93–106. MR2409717

McFadden, D. (1978). Modeling the choice of residential location. In *Spatial Interaction Theory and Planning Models* (A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull, eds.) 75–96. North Holland, Amsterdam.

Morales, C. F. (2005). Estimation of max-stable processes using Monte Carlo methods with applications to financial risk assessment. Ph.D. thesis, Dept. Statistics and Operations Research, Univ. North Carolina, Chapel Hill.

Morgan, B. J. T. (1983). Observations on quantit analysis. *Biometrics* **39** 879–886.

Roberts, S. (2000). Extreme value statistics for novelty detection in biomedical data processing. *Science, Measurement and Technology, IEE Proceedings* **147** 363–367.

Sang, H. and Gelfand, A. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* **16** 407–426.

Shmueli, G. and Koppius, O. (2009). The challenge of prediction in information systems research. Robert H. Smith School Research Paper No. RHS 06-058.

Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika* **72** 67–90. MR0790201

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Statist. Sci.* **4** 367–393. MR1041763

Smith, R. L. (2003). Statistics of extremes, with applications in environment, insurance and finance. In *Extreme Values in Finance, Telecommunications and the Environment* (B. Finkenstadt and H. Rootzen, eds.) 1–78. Chapman and Hall/CRC Press, London.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 583–639. MR1979380

Stavins, J. (2003). Perspective on payments: Electronic payments networks benefit banks, businesses, and consumers. Why do so few use them? *Regional Review* **13** 6–9.

Stukel, T. (1988). Generalized logistic models. *J. Amer. Statist. Assoc.* **83** 426–431. MR0971368

Thompson, M. L., Reynolds, J., Cox, L. H., Guttorp, P. and Sampson, P. D. (2001). A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric Environment* **35** 617–630.

Wang, X. (2010). Supplement to "Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption." DOI: 10.1214/10-AOAS354SUPP.

Wedel, M. and DeSarbo, W. (1993). A latent class binomial logit methodology for the analysis of paired comparison choice data: An application reinvestigating the determinants of perceived risk. *Decision Sciences* **24** 1157–1170.

Wu, Y., Chen, M.-H. and Dey, D. (2002). On the relationship between links for binary response data. *J. Stat. Stud. Special Volume in Honour of Professor Mir Masoom Ali's 65th Birthday* 159–172. MR1970193

National Institute
  of Statistical Sciences
19 T.W. Alexander Drive
Research Triangle Park,
  North Carolina 27709
USA
E-mail: xiawang@niss.org

Department of Statistics
University of Connecticut
215 Glenbrook Rd. U-4120
Storrs, Connecticut 06269
USA
E-mail: dipak.dey@uconn.edu